

---

# HER: A PLATFORM TO EXPLORE THE POSSIBILITY OF NATURAL CONVERSATIONAL AI

---

A PREPRINT

Steve J. Bickley, Ho Fai Chan, Bang Dao, Benno Torgler, Son Tran

Panalogy Lab  
Brisbane, 4000  
[steven.bickley@panalogy-lab.com](mailto:steven.bickley@panalogy-lab.com)

20.10.2023

## ABSTRACT

Our platform, “her”, aims to advance the rapidly evolving field of natural conversational Artificial Intelligence (AI) by integrating the emergent capabilities of Augmented Language Models (ALMs) with a dual voice and text modality. In essence, natural conversational AI is about designing AI systems to interact with humans in a way that feels “natural”. This includes understanding and generating human language, recognizing speech, interpreting context, resolving ambiguity, and even having an understanding and expression of emotions and empathy. The ultimate goal of natural conversational AI is to create artificial agents that can converse with a human just like another human would, making human-computer interactions as seamless and intuitive as possible, and opening up prospects for long-lived human-AI social interactions. Through “her”, we aim to comprehend the underlying principles and mechanisms that make AI interactions feel more natural and intuitive to humans. We are keen on dissecting the factors that contribute to the perceived “naturalness” of a conversation with AI and developing metrics to quantitatively evaluate this quality over time. Harnessing the power of state-of-the-art (SOTA) AI models, we have integrated advanced techniques and technologies into a unified platform capable of engaging in natural conversations. This innovative system will not only advance our understanding of AI-human interactions but also bring practical solutions to various sectors. Key applications for our research lie in the realms of (among others) education and entertainment. In education, this platform could serve as an intelligent tutor providing personalized assistance, while in entertainment, it could revolutionize how audiences interact with digital content. This project, therefore, holds substantial promise in fostering a more seamless and dynamic interaction between AI and human users, propelling us further into the age of natural conversational AI.

**Keywords** Natural Conversation • Social Science • Social AI • Large Language Models • Augmented Language Models • Alignment Research • Natural Conversation • Experimental Method

# 1 Motivation

We built “her” with the following motivations in mind:

- Language models such as GPT-4, Bard, Claude, etc. are touted as a pathway to AGI, and the ultimate test of AGI is the capability to make complex decisions involving interactions with humans and other artifacts in complex social settings (not just writing a poem or making a video). Making these kinds of complex social decisions, e.g. making friends, choosing spouses, pursuing a degree, trusting an AI system, telling a lie, etc., is what humans do every day in their societies.
- Thus, making complex social decisions is crucial test of AGI, and every social decision seems to start and end with a natural conversation. Natural conversations create decision premises that underpin decision processes. Thus, being able to engage in natural conversations plays a key role in designing AGI. Here, we define natural conversations are conversations that include, but not limited to, these elements: critical yet constructive and creative engagement, adaptive spontaneity, multi-modal and adaptive conversation style, evolving identity and perception formation, contextual understanding and expression of emotions, multi-level reflection, dynamic memories, grounded decision-making knowledge, and understanding of uncertainty.
- “her” is a platform to study and design AI system capable of engaging in natural conversations with humans. The platform was designed to explore to what extent we can address the limits of current LMs and their hard-coded prompt-engineering variants in terms of engaging in natural conversations. These models inherently lack most elements of naturalness when engaging in complex conversations. Our research explores ways to address this inherent limitation by extending the models with additional layers that learn to develop the elements of naturalness over time, through experience gained from past conversations.

# 2 Background

The development of Large Language Models (LLMs) marks a significant milestone for conversational AI. These models, which leverage sophisticated algorithms to understand and generate human-like text, have redefined our ability to create systems capable of engaging in meaningful dialogues. Particularly as they become increasingly convincing and capable through augmentations (i.e., Augmented Language Models or ALMs) via techniques like in-context learning and chain-of-thought prompting (Lee et al., n.d.), external memory, and the integration of other and multiple modalities (e.g., sound, visual). See Mialon et al. (2023) for a seminal review on ALMs. The increased complexity and fluency of these ALMs have brought us one step closer to the realization of more natural, human-like conversations with machines.

Yet, it’s important to differentiate between a conversation and a ‘natural’ conversation. While a conversation may involve a simple exchange of information (e.g., question-answering, a recommendation), a ‘natural’ conversation is characterized by adaptive spontaneity, critical and constructive engagement, context-awareness, emotional understanding, contextual memories, grounded knowledge, and the ability to manage the flow of dialogue, and often intertwined with complex elements such as humor, idioms, cultural nuances, and ambiguity.

Regarding conversational agents, findings from certain human-computer interaction (HCI) studies indicate that humans often prioritize the transactional aspects of conversation, such as task completion and goal attainment, over the more social, interactive elements (Clark et al., 2019). This mirrors the conversational requirements with strangers or casual acquaintances. However, it’s worth noting that such a focus may shift based on factors like user characteristics (Riefe & Benz, 2021) and the specific context of use (Rheu et al., 2021). Intriguingly, marketing research presents a more favorable view regarding the benefits of more natural conversational agents (Mariani et al., 2023). This optimism could suggest the significant value that consumers place on conversational AI now and into the future.

Clark et al. (2019) posed an intriguing question: what makes a good conversation? According to their study, humans value mutual understanding, trustworthiness, active listening, and humor in their conversations, but these traits’ desirability can differ across various types of human-human relationships. When it comes to human-agent interactions, these tend to be more transactional, utilitarian, and one-sided. However, sensitivity to the context of interaction, understanding of the conversation type required, and the underlying interaction purpose remain vital (p. 475). It is also worth noting that social functions of conversation play a critical role in establishing common ground, trust, and a sense of connection among participants. So, both transitional and social talk are important, and whilst serving distinct purposes, they frequently intersect in the context of natural conversations.

Why does natural conversation matter? The more ‘natural’ an AI’s conversational capabilities are, the more intuitive, engaging, and satisfying the user’s experience becomes. This seamless interaction fosters trust and understanding, reduces frustration, and increases the chances of successful communication.

So, why AI? AI's unique attributes make it an ideal candidate for facilitating natural conversation. Its limitless patience enables it to handle complex interactions without exhaustion. Its vast knowledge base allows it to provide accurate information across an array of topics. Its capability to simulate compassion can make conversations more comforting. Its constant availability ensures round-the-clock assistance, and its universal application means it can be customized to serve diverse needs across different cultures, languages, and domains.

At Panalogy Lab, we recognise that we are at the precipice of a new era where AI systems can function as enduring agents capable of making complex decisions, while interacting with humans and other entities in real-world social contexts. We acknowledge the monumental challenge of aligning the decision-making behaviors of these social AI systems with the evolving intentions and values of human society, an area whose complexity we are still endeavoring to fully comprehend. Our approach, however, is firmly rooted in this challenge. We believe that by infusing AI systems with social and emotional intelligence, we can create agents that can more effectively understand and mimic human conversational patterns. Through this, we aim to revolutionize the way humans interact with technology, pushing the boundaries of conversational AI and striving towards a future where AI not only understands our language but also the complex social fabric in which it is embedded.

### 3 Platform Design and Features

The platform's design process is guided by the following principles.

**Scientific focus on natural conversation:** the platform should be able to create an authentic feeling of having face-to-face natural conversations. Here, we take a scientific approach to identifying key elements of what make a conversation natural by carefully reviewing existing ideas in relevant literature and creating new constructs based on our own insights. These elements include, but not limited to, these elements: critical yet constructive and creative engagement, adaptive spontaneity, multi-modal and adaptive conversation style, evolving perception formation, contextual understanding and expression of emotions, multi-level reflection, dynamic memories, grounded perception and decision-making knowledge, and understanding of uncertainty. With these elements identified, we collect conversational data and transformed them into parameters that can be used by scientific methods to evaluate the naturalness of a conversation.

**Evolving and grounded companionship:** we started with the premise that natural conversations, if sustained over time, may result in an emergent companionship between human users and the platform. This companionship, however, must satisfy two criteria. First, it must be a result of learning from conversational evolution rather than instructions embedded in rigidly engineered prompt. Second, the platform's operations should be grounded in the sense that LMs do not actually "understand" what a conversation means in the best possible way and that they should not be instructed to "pretend" to assume a human character. Prompt engineering is the most obvious evidence showing that LMs lacks the kind of nuance "understanding" that humans have: a slightly rephrased can result in a significant different outcome. Our goal is to see whether, with evolution based on context learning based on our scientific framework, a new kind of companionship between humans and AI systems can emerge.

**Ease of interaction :** The goal is to make users' interaction with the platform as intuitive and simple as possible while making conversations flow smoothly. Since computers and humans have different ways of initiating and engaging in a conversation, especially the use of sensory elements, achieving this goal is an on-going process. The best solution is likely a compromise between what can be made possible by existing software and hardware technologies and how much users find the platform useful.

Based on these principles, "her" was built with several components as shown in Figure 1.

The platform's components were designed to enhance user interaction and facilitate research.

**Frontend Interface:** The interface serves as the gateway for users to interact with the platform. It provides users with simple voice and text input together with basic parameter configurations and document upload. The interface is designed to be user-friendly and intuitive, minimizing the learning curve for new users.

**Databases:** user credential database was separated from per-user conversation data, enabling optimization of platform's operations. The conversation data includes on-the-fly records of each message exchange in each conversation, context learning data, and relevant data. One important feature of this data is a set of vectors that capture various context of each conversation, creating a memory foundation for the platform.

**Backend interface:** This is where a conversation request is started and ended. The backend interface is responsible for authenticating users, processing input data and sending relevant responses to the frontend listener. It executes these tasks through appropriate interface calls to the user credential database and the conversation component.

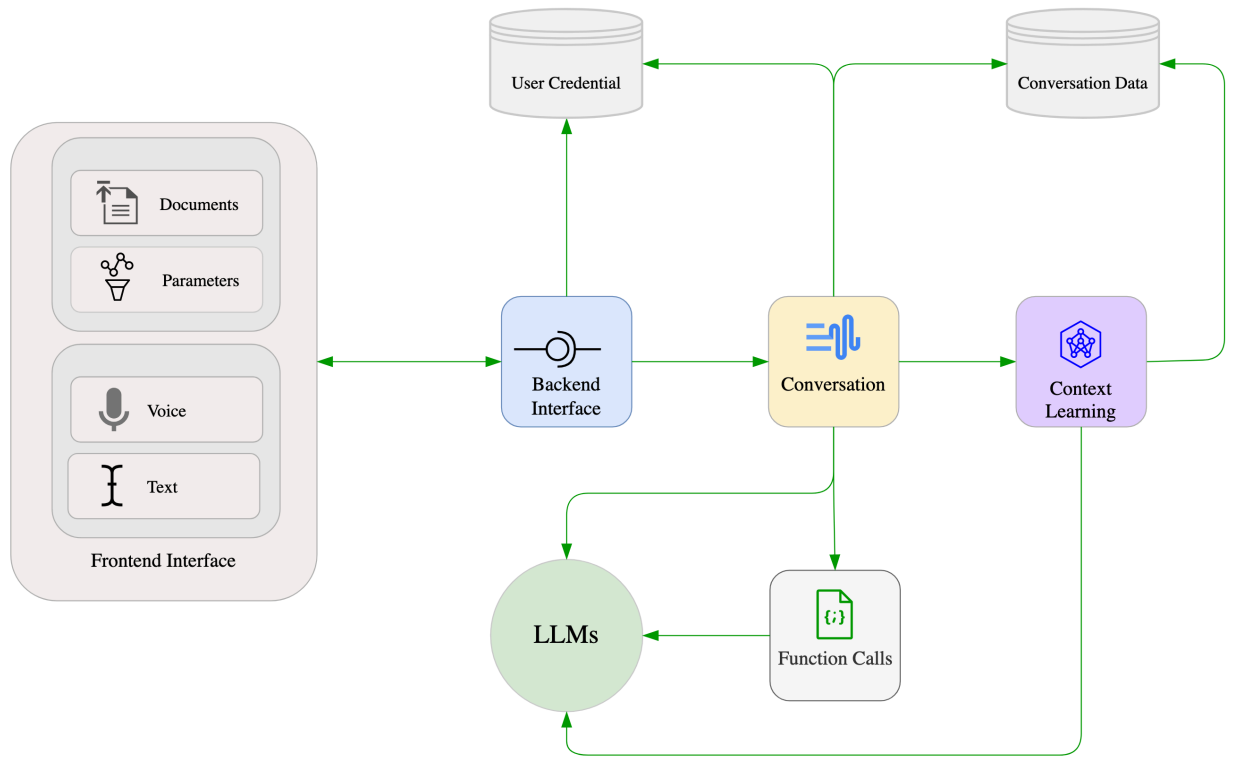


Figure 1: Key components of the platform design

**Conversation:** This is where a conversation process starts and ends. The conversation component is responsible for initializing per-user conversation context using historical conversation data, processing messages and document from users, performing conversation prompt parameterization, initiating function calls and context learning, and sending appropriate requests to LMs. An important feature of this component is the parameterization of prompts for LMs. Instead of using static prompts, the platform utilized parameterized prompts with the parameters being replaced with appropriate value and linguistic constructs learned from the latest conversation with the user. This on-the-fly prompt construction enables adaptive spontaneity and conversation style that usually lead to novel surprises. The conversation component serves as a conduit, ensuring efficient and correct data flow within the platform. It also internalizes the complexity of prompt engineering and related operations needed to ensure that the LLMs behave consistently in terms of the structure of their response.

**Context learning:** This component performs background learning tasks that enable the platform to observe conversations and extract insights from which future conversations will incorporate. These insights are gradually collated into layers of reflection enabling evolving perception of self and the world experienced through conversations. The insights will be incorporated into future conversations as on-the-fly parameters.

**Function calls:** Function calls are unique features of OpenAI's GPT class of models. They augment a LM with the capability to return a function call encoded in a JSON format that can be parsed to invoke customized functions that perform tasks that make the LM more factual and powerful. Besides providing LMs with a coherent way to generate grounded and up-to-date content unavailable in the stock models, function calls can be used, together with context learning and on-the-fly parameterization, to perform complex tasks such as contextual recall and learning to perceive human experience through conversations. While function calls are still an experimental feature and often limited to simple tasks dealing with simple data structure, our platform has been equipped with function calls that can perform complex tasks such as reflection in a reliable way.

As a result of our platform design choices, the following features have been developed:

- Automatic switching between languages is more reliable and natural, and this applies to both text and voice responses.
- ‘her’ can vary speaking rate and pitch to better express emotions that are adaptive to the conversational context.
- ‘her’ has both short-term and long-term memories to store conversational context.
- ‘her’ can gradually make observations about your character, enabling adaptive conversations. ‘her’ also makes observations of ‘her’ self. You can test this capability by asking “her” to tell you what ‘her’ thinks of you after chatting a while. The observations can be critical from time to time, and they are done in a way to minimize the impact on response speed.
- ‘her’ also gradually learn how to perceive the world through context learning and on-the-fly prompt parameterization.
- Gradually, ‘her’ will develop a sense of itself and a sense of your character in a conversational context. This is done by gradually observing . In other words, ‘her’ will develop a distinct personality and conversational style that are unique to you. This personality is a product of conversational evolution, not hard-coded character.
- Recalling memories is done via a function call. This function is triggered automatically when the GPT model detects contextual signal indicating that the user wants to talk about things that was or happened in the past.
- ‘her’ will reflect on or learn how to perceive an emotional experience when the user talks about emotions or feelings. This is also done automatically via function call.
- The user can upload a document and ‘her’ will read, analyze, and answer the user’s queries when there are appropriate signals. This is also done automatically via a function call whenever the user mentions document-related content. Otherwise, the document content will be ignored.
- The user can mention a website address, such as panalogy-lab.com, and ‘her’ will automatically create a query inferred from the user message, read the website content, and respond to the query. This is also done automatically via a function call.
- Conversation data is now network-ready for network construction and analysis. Moving forward, we plan to build a reinforcement learning agent operating on network measures capturing systematic meaning of naturalness. This will enable ‘her’ to be adaptive on another level.
- Overall, an interesting feature of this update is that ‘her’ will from time to time display emergent behaviors by using function calls in a ways never expected before.

## 4 Applications

With the emergent capabilities of Large Language Models (LLMs) and the ongoing development of Augmented Language Models (ALMs), understanding the underlying premises that govern their behaviors and decision-making processes becomes increasingly crucial, especially when they operate in complex social settings and have real-world impact. This understanding is essential to ensure the alignment of social AI with human values.

ALMs, which are essentially built on top of pre-trained LLMs like OpenAI’s gpt-4, incorporate various elements such as retrieval plug-ins, different learning techniques (few-shot), diverse prompting methods (such as chain-of-thought, self-model, and contextual prompts), functional coding, and integration with other modalities like voice, vision, and sound (Mialon et al., 2023). Additionally, future iterations of ALMs are expected to incorporate different AI techniques such as reinforcement learning and symbolic logic, enhancing their knowledge organisation, reasoning, and learning capabilities.

Researchers have recognized the potential of LLMs as valuable tools to study and probe the human mind and society (Arora et al., 2023; Binz & Schulz, 2023; Korinek, 2023; Miotto et al., 2022), given their training on vast amounts of human data and their ability to generate human-like text. Scholars have also discussed their potential in simulating human subjects (Aher et al., 2023; Horton, 2023; Park et al., 2022). Consequently, researchers from various disciplines such as behavioral economics, cognitive psychology, social psychology, linguistics have now started to investigate LLMs’ behavior and decision-making processes and its use as a scientific tool. However, the procedures and tuning of LLMs (e.g., temperature, context window, prompt context and structure) for judgment and evaluation of alignment are not yet standardized or consistently applied across studies. Moreover, digital literacy and programming skills continue to present significant obstacles for many researchers to implement such research robustly at scale, particularly those in behavioral economics and the social sciences.

Considering the fast-paced nature of research and development in AI at the moment, it is essential to also extend our focus beyond pre-trained LLMs and consider the emergent capabilities and value systems of ALMs within various

different social contexts. ALMs are increasingly augmented with additional tools and various prompting techniques, spanning different context windows and incorporating other modalities. Furthermore, these ALMs are now actively performing real-world actions. Calling a tool in the context of ALMs often involves having an impact on the virtual or physical world and observing the resulting effects, which are typically integrated into the ALM's ongoing context. Moreover, ALMs are increasingly engaging in delegate actions such as carrying out transactions on our behalf or responding to customer queries and emails in human-like ways.

By acknowledging the advancements in ALMs and the complex nature of their interactions with the world, we can gain a comprehensive understanding of the premises underlying their behaviors and decision-making processes by benchmarking through survey and experimental methods. This knowledge is crucial for ensuring the development of responsible and aligned social AI systems that reflect human values for the benefit of all humankind.

SurveyLM facilitates this exploration in an easy and intuitive manner. It empowers researchers to investigate the behaviors and decision-making of LLMs and ALMs in a robust and systematic way, using an easy-to-use, click-and-play online interface.

The SurveyLM platform is highly versatile and adaptable to a multitude of decision-making scenarios and experimental settings. Here are just some potential applications:

1. **Survey Data Generation:** Simulate agents to answer an array of survey questions, creating a rich dataset that can mimic diverse, human-like responses to these questions. This can be particularly useful in preliminary research phases, hypothesis testing, or for enhancing existing datasets.
2. **Allocation Games:** Simulate scenarios where agents need to make decisions about resource allocation. This could involve public goods games, bargaining games, prisoner's dilemma scenarios, or other economic games that explore cooperation, competition, and negotiation.
3. **Cognitive Psychology Experiments and (Multi-Stage) Scenarios:** Simulate cognitive tasks and adaptation processes to understand decision-making processes, memory, attention, perception, and problem-solving strategies.
4. **Social Interaction Simulations:** Model and simulate complex social interactions within groups, examining phenomena like group dynamics, communication patterns, social influence, and conformity.
5. **Consumer Behaviour Analysis:** Simulate buying decisions of agents to understand patterns in consumer behaviour, product preferences, and purchase rationales.
6. **Policy Impact Analysis:** Simulate reactions to new policies or regulations to gauge potential public response and impact (i.e., to understand emergent macro behaviours).
7. **Risk-taking (Multi-Stage) Scenarios:** Study decision-making under uncertainty or risk, such as in gambling or investment scenarios.
8. **Health-related Decision-making:** Explore choices related to health behaviours, preventative measures, treatment options, etc.
9. **Educational Settings:** Understand learning behaviour, knowledge acquisition, and responses to different teaching methods.
10. **Environmental Decision-making:** Simulate agent decisions about resource usage, conservation behaviours, and responses to environmental policies.

By simulating decision-making across a large spectrum of randomized agent demographic attributes (e.g., age, gender, education level, personality, etc.), SurveyLM provides a unique platform to investigate even the most sensitive, challenging, or otherwise taboo subjects/topics that are typically difficult to broach with human research participants, such as sexuality, drug use or life-event shocks. By probing these areas in a simulated environment, we leverage the potential of ALMs to explore sensitive topics (e.g., health, social, economic, ethical, etc) in a safe and ethical environment.

SurveyLM's potential is vast when it comes to exploring, e.g., sensitive, heated, challenging or taboo topics. Here are a few examples:

1. **Mental Health:** Explore attitudes and behaviours around mental health issues, which are often stigmatized or misunderstood.
2. **Addiction:** Understand the complex dynamics of substance use and addiction, and the social attitudes towards these subjects.
3. **Sexuality:** Explore attitudes towards various sexual orientations, gender identities, or sexual behaviours.

4. **Religion and Faith:** Assess how individuals interact with religious beliefs and practices, including perceptions of other religions.
5. **Political Extremism:** Investigate the drivers of extreme political views, intolerance, or radicalisation.
6. **Race and Ethnicity:** Examine attitudes towards different races and ethnicities, including instances of bias, discrimination, and prejudice.
7. **Immigration:** Assess perceptions and misconceptions about immigration and immigrants.
8. **Body Image:** Explore attitudes and pressures around body image and physical appearance.
9. **(Economic) Inequality:** Understand perspectives on wealth distribution, poverty, and economic disparity or inequalities in general.
10. **Climate Change and Natural Disasters:** Investigate attitudes and beliefs about climate change and natural disasters, environmental responsibility, and sustainability.

These topics are typically difficult to discuss openly, but SurveyLM provides a secure and confidential platform for exploring them in depth.

We are confident that as SurveyLM evolves and adapts over time, it will uncover numerous other promising areas of application. The examples provided here represent just a fraction of the platform’s potential, highlighting only the possibilities we, and others in the field, have identified to date. With the rapidly advancing landscape of social science research, there is no doubt that the scope of SurveyLM’s application will continue to expand, revealing even more groundbreaking possibilities in the future.

## 5 Future Developments

Despite its advanced design and capabilities, our research platform faces certain limitations that we are actively seeking to address. These hurdles present opportunities for enhancement and refinement, contributing to the platform’s ongoing evolution.

**Simplistic Agent Profile Configuration:** A significant limitation lies in the current mechanistic profile configuration for each agent (e.g., you are <AGE>, your personality is <BIG 5 PROFILE>, and reside in <LOCATION>). We do not draw on or attempt to simulate human subjects from demographic backgrounds of past survey respondents, as in e.g., (Argyle et al., 2023). Standard profile constructs often used in survey studies underpin this design, providing a simplified interaction model for users. However, these constructs’ rudimentary nature restricts the context within which the GPT models function, sometimes compromising the depth and richness of their responses. To capture the nuanced, multifaceted nature of human contexts, we need a more sophisticated approach. The solution we are developing and testing is a custom profile prompt feature, which will enable users to create intricate, context-sensitive profiles for their agents (e.g., profile construction by story telling). By broadening the contextual basis of agent profiles, we anticipate an enhancement in the model responses’ relevance and applicability.

**OpenAI API Rate Limit Constraints:** The rate limits imposed by OpenAI’s API can influence the stability of output and latency, creating potential bottlenecks for users requiring high-volume, real-time access to the models while being able to receive responses to their complex request in expected formats. A good solution depends on three things. First, users must be able to obtain API keys that have the desirable rate limits. Second, OpenAI, and other LLM providers, will increase or phase out rate limits. Third, the platform’s batching and request mechanisms must be robust. We are currently enhancing our concurrent request handler to optimize request scheduling and execution, thus maximizing throughput within rate limits for a given API keys. Additionally, we are in discussions with OpenAI to explore ways to improve throughput and latency.

**Model Diversity:** Our platform currently supports only OpenAI’s models, chosen for their leading-edge capabilities and robust API access. This model-specific dependency could limit the platform’s flexibility, as different models may offer unique strengths and capabilities that could be beneficial in diverse research scenarios. We are therefore actively testing other open source and commercial AI models to potentially integrate into our platform (Bai et al., 2022; Touvron et al., 2023), thus expanding its versatility and research applicability. It should be noted that we avoid models that essentially add new prompt-engineering layers on top of base ALMs to improve decision performance in specific tasks (Shinn et al., 2023; Yao et al., 2023).

**Realistic Condition Profiles:** A minor drawback of our platform is the occasional generation of unrealistic agent profiles due to the randomness of profile construction. This approach also means that sometimes we end up with “interesting” agent profile combinations that may seldom present in the real world (e.g., a male lesbian). While rare, these cases can disrupt the research process and lead to unrealistic model responses. One solution lies in conditional profile construction, where agent attributes are selected based on real-world prevalence and correlations. However, it

is important to note that this randomness can sometimes yield unique case studies that might not have been otherwise considered, offering unexpected insights and interesting research avenues.

**Multi-Agent Games and Interaction:** Currently, SurveyLM allows for the simulation of an agent's participation in games with other players only when the other players and interaction rules are hard-coded into the uploaded questions and answer instruction. However, we are yet to develop the capacity for interactions between different agents within the same simulated population. To achieve this, a series of sophisticated enhancements would be necessary. Key among these improvements is the ability to form and manage groups of agents. These groups function as independent social entities, engaging in intricate social interactions within their own set boundaries. In this envisioned application, users could define group sizes, roles within these groups (e.g., the proposer and responder roles in the ultimatum game), and any variations thereof. Furthermore, we would offer the ability to set randomisation parameters for both roles and variations, adding yet another layer of complexity and realism to the simulations. Another captivating idea under this future avenue is the introduction of a chat function between different agents. This feature would allow the observation of direct communication patterns and language use within and across agent groups, offering researchers another dimension to their social agent studies.

In summary, while our platform faces certain limitations, we view these as opportunities for growth and enhancement. Our commitment to continuous development and user satisfaction drives us to persistently explore innovative solutions to these challenges. This means that we are open for feedback and suggestions. As we progress on this journey, we look forward to unlocking further potential in facilitating complex research through advanced ALM models.

For beta access and further details about the SurveyLM platform, please contact the corresponding author of this paper.



## References

- Aher, G., Arriaga, R. I., & Kalai, A. T. (2023). *Using large language models to simulate multiple humans and replicate human subject studies* (arXiv:2208.10264). arXiv. <http://arxiv.org/abs/2208.10264>
- Argyle, L. P., Busby, E., Gubler, J., Bail, C., Howe, T., Rytting, C., & Wingate, D. (2023). *AI chat assistants can improve conversations about divisive topics* (arXiv:2302.07268). arXiv. <http://arxiv.org/abs/2302.07268>
- Arora, A., Kaffee, L.-A., & Augenstein, I. (2023). *Probing pre-trained language models for cross-cultural differences in values* (arXiv:2203.13722). arXiv. <http://arxiv.org/abs/2203.13722>
- Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., ... Kaplan, J. (2022). *Constitutional AI: Harmlessness from AI feedback* (arXiv:2212.08073). arXiv. <https://doi.org/10.48550/arXiv.2212.08073>
- Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6), e2218523120. <https://doi.org/10.1073/pnas.2218523120>
- Horton, J. J. (2023). *Large language models as simulated economic agents: What can we learn from homo silicus?* (arXiv:2301.07543). arXiv. <http://arxiv.org/abs/2301.07543>
- Korinek, A. (2023). *Language models and cognitive automation for economic research* (30957). National Bureau of Economic Research. <https://doi.org/10.3386/w30957>
- Mialon, G., Dessì, R., Lomeli, M., Nalmpantis, C., Pasunuru, R., Raileanu, R., Rozière, B., Schick, T., Dwivedi-Yu, J., Celikyilmaz, A., Grave, E., LeCun, Y., & Scialom, T. (2023). *Augmented language models: A survey* (arXiv:2302.07842). arXiv. <https://doi.org/10.48550/arXiv.2302.07842>
- Miotto, M., Rossberg, N., & Kleinberg, B. (2022). *Who is GPT-3? An exploration of personality, values and demographics* (arXiv:2209.14338). arXiv. <http://arxiv.org/abs/2209.14338>
- Park, J. S., Popowski, L., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2022). *Social simulacra: Creating populated prototypes for social computing systems* (arXiv:2208.04024). arXiv. <http://arxiv.org/abs/2208.04024>
- Shinn, N., Cassano, F., Labash, B., Gopinath, A., Narasimhan, K., & Yao, S. (2023). *Reflexion: Language agents with verbal reinforcement learning* (arXiv:2303.11366). arXiv. <https://doi.org/10.48550/arXiv.2303.11366>
- Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., ... Scialom, T. (2023). *Llama 2: Open foundation and fine-tuned chat models* (arXiv:2307.09288). arXiv. <https://doi.org/10.48550/arXiv.2307.09288>
- Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). *ReAct: Synergizing reasoning and acting in language models* (arXiv:2210.03629). arXiv. <https://doi.org/10.48550/arXiv.2210.03629>